

KB



PCT09

ENTERED

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/914,543

DATE: 04/25/2002

TIME: 11:38:44

Input Set : A:\DIVER1150-5.TXT

Output Set: N:\CRF3\04252002\I914543.raw

#9

```

4 <110> APPLICANT: DIVERSA CORPORATION
5     LAM, David
6     MATHUR, Eric
8 <120> TITLE OF INVENTION: ENDOGLUCANASES
11 <130> FILE REFERENCE: DIVER1150-5
13 <140> CURRENT APPLICATION NUMBER: 09/914,543
14 <141> CURRENT FILING DATE: 1998-11-22
16 <150> PRIOR APPLICATION NUMBER: PCT/US97/08793
17 <151> PRIOR FILING DATE: 1997-05-22
19 <150> PRIOR APPLICATION NUMBER: US 08/651,572
20 <151> PRIOR FILING DATE: 1996-05-22
22 <160> NUMBER OF SEQ ID NOS: 52
24 <170> SOFTWARE: FastSEQ for Windows Version 4.0
26 <210> SEQ ID NO: 1
27 <211> LENGTH: 2529
28 <212> TYPE: DNA
29 <213> ORGANISM: Thermotoga maritima
31 <220> FEATURE:
32 <221> NAME/KEY: CDS
33 <222> LOCATION: (1)...(2523)
34 <223> OTHER INFORMATION: AEPII 1a Archaeal Endoglucanase
36 <400> SEQUENCE: 1
37 atg ata aac gtt gca acg gga gag gag acc cca ata cac ctc ttt gga      48
38 Met Ile Asn Val Ala Thr Gly Glu Glu Thr Pro Ile His Leu Phe Gly
39 1          5          10          15
41 gtc aac tgg ttc ggc ttt gag aca ccg aac tac gtt gtt cac ggc cta      96
42 Val Asn Trp Phe Gly Phe Glu Thr Pro Asn Tyr Val Val His Gly Leu
43          20          25          30
45 tgg agt agg aac tgg gag gac atg ctc ctc cag atc aag agc ctt ggc      144
46 Trp Ser Arg Asn Trp Glu Asp Met Leu Leu Gln Ile Lys Ser Leu Gly
47          35          40          45
49 ttc aat gcg ata agg ctt ccc ttc tgt acc cag tca gta aaa ccg ggg      192
50 Phe Asn Ala Ile Arg Leu Pro Phe Cys Thr Gln Ser Val Lys Pro Gly
51          50          55          60
53 acg atg cca acg gcg att gac tac gcc aag aac cca gac ctc cag ggt      240
54 Thr Met Pro Thr Ala Ile Asp Tyr Ala Lys Asn Pro Asp Leu Gln Gly
55 65          70          75          80
57 ctt gac agc gtc cag ata atg gag aaa ata atc aag aag gct gga gac      288
58 Leu Asp Ser Val Gln Ile Met Glu Lys Ile Ile Lys Lys Ala Gly Asp
59          85          90          95
61 ctg ggc ata ttc gtg ctc ctc gac tac cac aga ata gga tgc aac ttc      336
62 Leu Gly Ile Phe Val Leu Leu Asp Tyr His Arg Ile Gly Cys Asn Phe
63          100          105          110

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/914,543

DATE: 04/25/2002

TIME: 11:38:44

Input Set : A:\DIVER1150-5.TXT

Output Set: N:\CRF3\04252002\I914543.raw

65	ata	gaa	ccc	cta	tgg	tac	acc	gac	agc	ttc	tcg	gag	cag	gac	tac	ata	384
66	Ile	Glu	Pro	Leu	Trp	Tyr	Thr	Asp	Ser	Phe	Ser	Glu	Gln	Asp	Tyr	Ile	
67			115					120					125				
69	aac	acc	tgg	gtt	gaa	gtc	gcc	cag	agg	ttc	ggc	aag	tac	tgg	aac	gtt	432
70	Asn	Thr	Trp	Val	Glu	Val	Ala	Gln	Arg	Phe	Gly	Lys	Tyr	Trp	Asn	Val	
71			130					135					140				
73	atc	ggc	gcg	gac	ctg	aag	aac	gaa	ccc	cac	agc	tca	agc	ccc	gca	cct	480
74	Ile	Gly	Ala	Asp	Leu	Lys	Asn	Glu	Pro	His	Ser	Ser	Ser	Pro	Ala	Pro	
75	145					150					155					160	
77	gcc	gcc	tac	act	gac	gga	agt	ggg	gcc	acg	tgg	gga	atg	ggc	aac	aac	528
78	Ala	Ala	Tyr	Thr	Asp	Gly	Ser	Gly	Ala	Thr	Trp	Gly	Met	Gly	Asn	Asn	
79					165					170					175		
81	gcc	acc	gac	tgg	aac	ctg	gcg	gct	gag	agg	ata	gga	agg	gca	att	ctg	576
82	Ala	Thr	Asp	Trp	Asn	Leu	Ala	Ala	Glu	Arg	Ile	Gly	Arg	Ala	Ile	Leu	
83				180					185					190			
85	gag	gtt	gcc	cca	caa	tgg	gtt	ata	ttt	gtt	gag	gga	acc	cag	ttc	acc	624
86	Glu	Val	Ala	Pro	Gln	Trp	Val	Ile	Phe	Val	Glu	Gly	Thr	Gln	Phe	Thr	
87			195					200					205				
89	acc	ccc	gag	ata	gac	ggg	agg	tac	aag	tgg	ggc	cac	aac	gcc	tgg	tgg	672
90	Thr	Pro	Glu	Ile	Asp	Gly	Arg	Tyr	Lys	Trp	Gly	His	Asn	Ala	Trp	Trp	
91			210				215					220					
93	ggc	gga	aac	ctt	atg	ggg	gtt	agg	aag	tac	cca	gtt	aac	ctg	ccc	agg	720
94	Gly	Gly	Asn	Leu	Met	Gly	Val	Arg	Lys	Tyr	Pro	Val	Asn	Leu	Pro	Arg	
95	225				230					235					240		
97	gac	aag	gtt	gtt	tac	agc	ccc	caa	gtt	tac	ggg	tca	gaa	gtt	tac	gac	768
98	Asp	Lys	Val	Val	Tyr	Ser	Pro	Gln	Val	Tyr	Gly	Ser	Glu	Val	Tyr	Asp	
99			245					250					255				
101	cag	ccc	tac	ttt	gac	ccc	ggg	gag	ggg	ttc	ccc	gac	aac	ctc	ccc	gaa	816
102	Gln	Pro	Tyr	Phe	Asp	Pro	Gly	Glu	Gly	Phe	Pro	Asp	Asn	Leu	Pro	Glu	
103			260					265					270				
105	ata	tgg	tac	cac	cac	ttc	ggc	tac	gta	aag	ctt	gat	ctc	ggg	tac	cct	864
106	Ile	Trp	Tyr	His	His	Phe	Gly	Tyr	Val	Lys	Leu	Asp	Leu	Gly	Tyr	Pro	
107			275					280					285				
109	gtt	gtt	ata	ggg	gag	ttc	gga	ggc	aag	tac	ggc	cat	ggg	gga	gac	ccg	912
110	Val	Val	Ile	Gly	Glu	Phe	Gly	Gly	Lys	Tyr	Gly	His	Gly	Gly	Asp	Pro	
111			290				295					300					
113	agg	gat	gtc	act	tgg	cag	aac	aag	ata	ata	gac	tgg	atg	atc	cag	aac	960
114	Arg	Asp	Val	Thr	Trp	Gln	Asn	Lys	Ile	Ile	Asp	Trp	Met	Ile	Gln	Asn	
115	305					310					315				320		
117	aaa	ttc	tgt	gac	ttc	ttc	tac	tgg	agc	tgg	aac	cca	aac	agc	ggg	gac	1008
118	Lys	Phe	Cys	Asp	Phe	Phe	Tyr	Trp	Ser	Trp	Asn	Pro	Asn	Ser	Gly	Asp	
119				325					330					335			
121	acc	ggg	gga	att	ctg	aag	gat	gac	tgg	acg	aca	ata	tgg	gag	gac	aag	1056
122	Thr	Gly	Gly	Ile	Leu	Lys	Asp	Asp	Trp	Thr	Thr	Ile	Trp	Glu	Asp	Lys	
123				340					345				350				
125	tac	aac	aac	ctg	aag	agg	ctc	atg	gac	agc	tgt	tct	gga	aac	gcc	act	1104
126	Tyr	Asn	Asn	Leu	Lys	Arg	Leu	Met	Asp	Ser	Cys	Ser	Gly	Asn	Ala	Thr	
127			355					360					365				
129	gcc	ccg	tcc	gtc	ccc	acg	aca	act	aca	aca	aca	agc	aca	ccg	cca	acg	1152

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/914,543

DATE: 04/25/2002

TIME: 11:38:44

Input Set : A:\DIVER1150-5.TXT

Output Set: N:\CRF3\04252002\I914543.raw

130	Ala	Pro	Ser	Val	Pro	Thr	Thr	Thr	Thr	Thr	Ser	Thr	Pro	Pro	Thr	
131		370				375					380					
133	acc	aca	acg	act	aca	aca	tcc	act	cca	acg	acc	act	acc	cag	acc	ccg
134	Thr	Thr	Thr	Thr	Thr	Thr	Ser	Thr	Pro	Thr	Thr	Thr	Thr	Gln	Thr	Pro
135	385					390					395					400
137	acc	acc	act	act	cca	act	acg	aca	acc	acc	acg	acc	aca	act	cct	tca
138	Thr	Thr	Thr	Thr	Pro	Thr	Thr	Thr	Thr	Thr	Thr	Thr	Thr	Thr	Pro	Ser
139					405						410					415
141	aat	aac	gtc	cca	ttt	gaa	att	gtg	aac	gtt	ctc	ccg	act	agc	tcc	cag
142	Asn	Asn	Val	Pro	Phe	Glu	Ile	Val	Asn	Val	Leu	Pro	Thr	Ser	Ser	Gln
143					420					425					430	
145	tac	gag	gga	acc	agc	gtg	gag	gtt	gta	tgt	gat	gga	acc	cag	tgt	gcc
146	Tyr	Glu	Gly	Thr	Ser	Val	Glu	Val	Val	Cys	Asp	Gly	Thr	Gln	Cys	Ala
147			435					440						445		
149	tcc	agc	gtt	tgg	gga	gct	ccg	aac	ctc	tgg	gga	gtc	gtt	aaa	atc	gga
150	Ser	Ser	Val	Trp	Gly	Ala	Pro	Asn	Leu	Trp	Gly	Val	Val	Lys	Ile	Gly
151			450				455							460		
153	aac	gcc	acc	atg	gac	ccc	aac	gtt	tgg	ggc	tgg	gag	gac	gtt	tac	aag
154	Asn	Ala	Thr	Met	Asp	Pro	Asn	Val	Trp	Gly	Trp	Glu	Asp	Val	Tyr	Lys
155	465					470					475					480
157	act	gca	ccc	cag	gac	att	gga	acc	ggc	agc	aca	aag	atg	gag	ata	agg
158	Thr	Ala	Pro	Gln	Asp	Ile	Gly	Thr	Gly	Ser	Thr	Lys	Met	Glu	Ile	Arg
159					485					490					495	
161	aac	ggg	gtg	ctc	aag	gtt	aca	aac	ctc	tgg	aac	atc	aac	atg	cat	ccg
162	Asn	Gly	Val	Leu	Lys	Val	Thr	Asn	Leu	Trp	Asn	Ile	Asn	Met	His	Pro
163				500					505					510		
165	aag	tat	aac	aca	atg	gca	tac	ccg	gag	gtc	ata	tac	ggc	gcc	aag	cct
166	Lys	Tyr	Asn	Thr	Met	Ala	Tyr	Pro	Glu	Val	Ile	Tyr	Gly	Ala	Lys	Pro
167			515					520						525		
169	tgg	ggc	aac	cag	cca	ata	aac	gct	ccg	aac	ttc	gtg	ctc	ccg	ata	aag
170	Trp	Gly	Asn	Gln	Pro	Ile	Asn	Ala	Pro	Asn	Phe	Val	Leu	Pro	Ile	Lys
171			530				535					540				
173	gtc	tcc	cag	ctt	ccg	agg	ata	ctc	gtt	gac	aca	aag	tac	acg	ctc	gaa
174	Val	Ser	Gln	Leu	Pro	Arg	Ile	Leu	Val	Asp	Thr	Lys	Tyr	Thr	Leu	Glu
175	545					550					555					560
177	aag	agc	ttc	ccg	gga	aac	aac	ttc	gcc	ttt	gag	gcc	tgg	ctc	ttc	aag
178	Lys	Ser	Phe	Pro	Gly	Asn	Asn	Phe	Ala	Phe	Glu	Ala	Trp	Leu	Phe	Lys
179					565				570						575	
181	gat	gcc	aac	aac	atg	agg	gca	cca	ggc	cag	ggg	gac	tac	gag	ata	atg
182	Asp	Ala	Asn	Asn	Met	Arg	Ala	Pro	Gly	Gln	Gly	Asp	Tyr	Glu	Ile	Met
183				580					585					590		
185	gta	cag	ctc	tac	atc	gag	ggc	ggc	tat	cct	gcg	ggc	tac	gac	aag	ggg
186	Val	Gln	Leu	Tyr	Ile	Glu	Gly	Gly	Tyr	Pro	Ala	Gly	Tyr	Asp	Lys	Gly
187			595					600					605			
189	cca	gtt	ctc	acc	gtt	gat	gtt	ccg	ata	atc	gtc	gat	gga	agg	ctt	gta
190	Pro	Val	Leu	Thr	Val	Asp	Val	Pro	Ile	Ile	Val	Asp	Gly	Arg	Leu	Val
191			610					615					620			
193	aac	cag	act	ttt	gag	ctc	tac	gac	gtc	ata	gcg	gat	gcc	gga	tgg	agg
194	Asn	Gln	Thr	Phe	Glu	Leu	Tyr	Asp	Val	Ile	Ala	Asp	Ala	Gly	Trp	Arg

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/914,543

DATE: 04/25/2002

TIME: 11:38:44

Input Set : A:\DIVER1150-5.TXT

Output Set: N:\CRF3\04252002\I914543.raw

```

195 625
197 ttc ttc acc ttc aag cca act aag aac tac aac ggc tca gag gtt gtg
198 Phe Phe Thr Phe Lys Pro Thr Lys Asn Tyr Asn Gly Ser Glu Val Val 1968
199
201 ttc gac tac acc aaa ttc ata gaa ata gtt gac aac tac ctc ggc ggt
202 Phe Asp Tyr Thr Lys Phe Ile Glu Ile Val Asp Asn Tyr Leu Gly Gly 2016
203
205 ggc agc ctc acg aac cac tac ctg atg tcc ctg gaa ttc ggt acc gag
206 Gly Ser Leu Thr Asn His Tyr Leu Met Ser Leu Glu Phe Gly Thr Glu 2064
207
209 ata tac acc aac ggg tgc acc tca ttc cca tgc aca gtg gac gta agg
210 Ile Tyr Thr Asn Gly Cys Thr Ser Phe Pro Cys Thr Val Asp Val Arg 2112
211
213 tgg acc ctt gac aag tac agg ttc atc ctg gcc cca gga aca atg gcc
214 Trp Thr Leu Asp Lys Tyr Arg Phe Ile Leu Ala Pro Gly Thr Met Ala 2160
215 705
217 act gag gag gcc atg aga gtt ctc gtc gga gag gtc cag cct ccc gct
218 Thr Glu Glu Ala Met Arg Val Leu Val Gly Glu Val Gln Pro Pro Ala 2208
219
221 tcc aca aca aca tcg cag acg act act tca acc aca acc cca acg ccc
222 Ser Thr Thr Thr Ser Gln Thr Thr Thr Ser Thr Thr Thr Pro Thr Pro 2256
223
225 act acc act act acg act cag act tca acc acc act aca acc acc tca
226 Thr Thr Thr Thr Thr Gln Thr Ser Thr Thr Thr Thr Thr Thr Ser 2304
227
229 ccg ccg aca acc acc gca cct gct cag gac gta att aag ctc agg tac
230 Pro Pro Thr Thr Thr Ala Pro Ala Gln Asp Val Ile Lys Leu Arg Tyr 2352
231
233 ccg gac gat ggg cag tgg ccc gag gcc cca att gac agg gat gga gac
234 Pro Asp Asp Gly Gln Trp Pro Glu Ala Pro Ile Asp Arg Asp Gly Asp 2400
235 785
237 gga aac cca gag ttc tac ata gaa ata aac ccg tgg aac ata ctg agc
238 Gly Asn Pro Glu Phe Tyr Ile Glu Ile Asn Pro Trp Asn Ile Leu Ser 2448
239
241 gct gaa agc tac gcc gag atg acc tac aac ttg agc agc ggg gtt ctc
242 Ala Glu Ser Tyr Ala Glu Met Thr Tyr Asn Leu Ser Ser Gly Val Leu 2496
243
245 cac tac gtc cag gcc ctg gat agt ata tgatga
246 His Tyr Val Gln Ala Leu Asp Ser Ile 2529
247
250 <210> SEQ ID NO: 2
251 <211> LENGTH: 841
252 <212> TYPE: PRT
253 <213> ORGANISM: Thermotoga maritima
255 <400> SEQUENCE: 2
256 Met Ile Asn Val Ala Thr Gly Glu Glu Thr Pro Ile His Leu Phe Gly
257 1 5 10 15
258 Val Asn Trp Phe Gly Phe Glu Thr Pro Asn Tyr Val Val His Gly Leu
259 20 25 30

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/914,543

DATE: 04/25/2002

TIME: 11:38:44

Input Set : A:\DIVER1150-5.TXT

Output Set: N:\CRF3\04252002\I914543.raw

```

260 Trp Ser Arg Asn Trp Glu Asp Met Leu Leu Gln Ile Lys Ser Leu Gly
261      35      40      45
262 Phe Asn Ala Ile Arg Leu Pro Phe Cys Thr Gln Ser Val Lys Pro Gly
263      50      55      60
264 Thr Met Pro Thr Ala Ile Asp Tyr Ala Lys Asn Pro Asp Leu Gln Gly
265 65      70      75      80
266 Leu Asp Ser Val Gln Ile Met Glu Lys Ile Ile Lys Lys Ala Gly Asp
267      85      90      95
268 Leu Gly Ile Phe Val Leu Leu Asp Tyr His Arg Ile Gly Cys Asn Phe
269      100      105      110
270 Ile Glu Pro Leu Trp Tyr Thr Asp Ser Phe Ser Glu Gln Asp Tyr Ile
271      115      120      125
272 Asn Thr Trp Val Glu Val Ala Gln Arg Phe Gly Lys Tyr Trp Asn Val
273      130      135      140
274 Ile Gly Ala Asp Leu Lys Asn Glu Pro His Ser Ser Ser Pro Ala Pro
275 145      150      155      160
276 Ala Ala Tyr Thr Asp Gly Ser Gly Ala Thr Trp Gly Met Gly Asn Asn
277      165      170      175
278 Ala Thr Asp Trp Asn Leu Ala Ala Glu Arg Ile Gly Arg Ala Ile Leu
279      180      185      190
280 Glu Val Ala Pro Gln Trp Val Ile Phe Val Glu Gly Thr Gln Phe Thr
281      195      200      205
282 Thr Pro Glu Ile Asp Gly Arg Tyr Lys Trp Gly His Asn Ala Trp Trp
283      210      215      220
284 Gly Gly Asn Leu Met Gly Val Arg Lys Tyr Pro Val Asn Leu Pro Arg
285 225      230      235      240
286 Asp Lys Val Val Tyr Ser Pro Gln Val Tyr Gly Ser Glu Val Tyr Asp
287      245      250      255
288 Gln Pro Tyr Phe Asp Pro Gly Glu Gly Phe Pro Asp Asn Leu Pro Glu
289      260      265      270
290 Ile Trp Tyr His His Phe Gly Tyr Val Lys Leu Asp Leu Gly Tyr Pro
291      275      280      285
292 Val Val Ile Gly Glu Phe Gly Gly Lys Tyr Gly His Gly Gly Asp Pro
293      290      295      300
294 Arg Asp Val Thr Trp Gln Asn Lys Ile Ile Asp Trp Met Ile Gln Asn
295 305      310      315      320
296 Lys Phe Cys Asp Phe Phe Tyr Trp Ser Trp Asn Pro Asn Ser Gly Asp
297      325      330      335
298 Thr Gly Gly Ile Leu Lys Asp Asp Trp Thr Thr Ile Trp Glu Asp Lys
299      340      345      350
300 Tyr Asn Asn Leu Lys Arg Leu Met Asp Ser Cys Ser Gly Asn Ala Thr
301      355      360      365
302 Ala Pro Ser Val Pro Thr Thr Thr Thr Thr Thr Ser Thr Pro Pro Thr
303      370      375      380
304 Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Gln Thr Pro
305 385      390      395      400
306 Thr Thr Thr Thr Pro Thr Thr Thr Thr Thr Thr Thr Thr Pro Ser
307      405      410      415
308 Asn Asn Val Pro Phe Glu Ile Val Asn Val Leu Pro Thr Ser Ser Gln

```

## VERIFICATION SUMMARY

PATENT APPLICATION: US/09/914,543

DATE: 04/25/2002

TIME: 11:38:45

Input Set : A:\DIVER1150-5.TXT

Output Set: N:\CRF3\04252002\I914543.raw

L:371 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!  
L:374 M:258 W: Mandatory Feature missing, <220> not found for SEQ ID#:3  
L:2276 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!  
L:2279 M:258 W: Mandatory Feature missing, <220> not found for SEQ ID#:17  
L:2335 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!  
L:2338 M:258 W: Mandatory Feature missing, <220> not found for SEQ ID#:19  
L:2544 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!  
L:2547 M:258 W: Mandatory Feature missing, <220> not found for SEQ ID#:21  
L:2753 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!  
L:2756 M:258 W: Mandatory Feature missing, <220> not found for SEQ ID#:23  
L:2972 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!  
L:2975 M:258 W: Mandatory Feature missing, <220> not found for SEQ ID#:25  
L:3188 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!  
L:3191 M:258 W: Mandatory Feature missing, <220> not found for SEQ ID#:27  
L:3409 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!  
L:3412 M:258 W: Mandatory Feature missing, <220> not found for SEQ ID#:29  
L:3672 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!  
L:3675 M:258 W: Mandatory Feature missing, <220> not found for SEQ ID#:31